Method

Infogroup's U.S. Business Databases is organized according to the U.S. Government's Standard Industrial Classification (SIC) Coding system that categorizes businesses by general business activity. Maintained are over 17,000 business category titles which net into 10,500 unique 6-digit (typical is 4-digits) SIC codes. Whereas the SIC system classified businesses based on the products or services they *produced*, NAICS classify businesses based on the *processes* used to create the products or services. Infogroup has added two additional digits to the 6-digit NAICS codes to better match conversion from 6-digit SIC codes, offering a significant improvement in granularity.

Infogroup vs. Census

It is not at all uncommon to see variance, and sometimes a significant variance, between aggregate industry verticals when looking at Infogroup data versus Census. There are several reasons for this:

- With Census, businesses are inputting what they believe to be the most appropriate industry code. However more often than not there are multiple codes that are similar and the business may not select what is actually the most appropriate as industry classification is not typically their strong suit.
- Infogroup uses Yellow Page headings as a guide to assign what is actually the most appropriate industry code, and telephone verify the businesses to confirm the accuracy. This helps to ensure the correct code is selected. As a result, by viewing primary and secondary codes, we typically see approximately 90% or better in terms of industry classification accuracy.
- Historic data sets will contain only records where Infogroup was able to validate the business prior to that years data cut. Meaning there would have been more records that may have fit the industry classification but were not in the verified file.
- Since a system is used that is SIC to NAICS based it would expand the opportunity for discrepancy. Meaning Infogroup may have created a subcategory that doesn't fit neatly into the 4-digit codes business owners/Census would present (again most business owners are not experts in Industry code selection).
- A BDRC comparison between YTS data and Census data show high number of establishments cluster in certain 6-digit NAICS which can significantly inflate or deflate particular NAICS categories, predominately due to certain industry segments that house financial, real estate, and tax reporting entities.

Secondary Classifications and YTS Verification

Infogroup does not assert any measure of quality for secondary industry classifications. Primary and secondary SIC and primary NAICS (no secondary NAICS as of this writing) are part of a perennial process of verification in the YTS database using both the Infogroup verified and pre-verified files applied to all years of the database where necessary.

August 3, 2017   bdrc.uwex.edu